

ESTIMATION NON PARAMÉTRIQUE OPTIMALE DU TAUX DE SAUT D'UN PROCESSUS MARKOVIEEN DÉTERMINISTE PAR MORCEAUX

Romain Azaïs & Aurélie Muller-Gueudin

*Inria Nancy-Grand Est, Team BIGS, Institut Élie Cartan de Lorraine,
Vandoeuvre-lès-Nancy, France
romain.azais@inria.fr & aurelie.gueudin@univ-lorraine.fr*

Résumé. Un processus markovien déterministe par morceaux est un processus stochastique dont la trajectoire est décrite par une équation différentielle perturbée par des sauts aléatoires en des instants aléatoires. Nous nous intéressons à l'estimation du taux de saut d'un tel processus observé en temps long sous une hypothèse d'ergodicité. Nous introduisons une classe d'estimateurs non paramétriques consistants et asymptotiquement gaussiens. Nous proposons de choisir l'estimateur de variance minimale, variance qui est elle-même à estimer.

Mots-clés. taux de saut, méthode à noyau, estimation non paramétrique, processus markovien déterministe par morceaux

Abstract. A piecewise-deterministic Markov process is a stochastic process whose behavior is governed by an ordinary differential equation punctuated by random jumps occurring at random times. We focus on the nonparametric estimation problem of the jump rate for such a stochastic model observed within a long time interval under an ergodicity condition. We introduce an uncountable class of kernel estimates of the jump rate and we establish their strong pointwise consistency as well as their asymptotic normality. We propose to choose among this class the estimator with the minimal variance, which is unfortunately unknown and thus remains to be estimated.

Keywords. jump rate, kernel method, nonparametric estimation, piecewise-deterministic Markov process

1 Introduction

Les processus markoviens déterministes par morceaux (PDMP pour l'anglais Piecewise-Deterministic Markov Processes) ont été introduits par Davis [5] dans les années 1980 comme une large classe de modèles stochastiques à temps continu non diffusifs. Ils apparaissent dans de nombreuses applications, notamment en biologie [2, 4, 8, 9], en neurosciences [7], et en fiabilité [3, 6]. Dès lors se pose la question de l'estimation des paramètres d'un tel processus à partir d'une ou plusieurs observations.

La trajectoire d'un PDMP (X_t) sur un ouvert E de $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est définie à partir de ses trois caractéristiques locales (λ, Q, Φ) :

- $\Phi : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}^d$ est le flot déterministe ;
- $\lambda : \mathbb{R}^d \rightarrow \mathbb{R}_+$ est le taux de saut ;
- $Q : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \rightarrow [0, 1]$ est le noyau de transition.

Nous introduisons les temps de sortie de E pour le flot et le flot parcouru à rebours : pour tout $x \in E$,

$$t^+(x) = \inf \{t > 0 : \Phi(x, t) \in \partial E\} \quad \text{et} \quad t^-(x) = \inf \{t > 0 : \Phi(x, -t) \in \partial E\}.$$

Nous supposons que ces caractéristiques satisfont les conditions usuelles [5, (24.8)]. Partant d'une condition initiale $X_0 = x$, la trajectoire de (X_t) est décrite comme suit. Le premier temps de saut est défini par sa fonction de survie,

$$\forall t \geq 0, \mathbb{P}(T_1 > t \mid X_0 = x) = \exp \left(- \int_0^t \lambda \circ \Phi(x, s) \, ds \right) \mathbb{I}_{\{t < t^+(x)\}}.$$

La trajectoire entre 0 et le premier temps de saut T_1 est donnée par,

$$\forall 0 \leq t \leq T_1, X_t = \begin{cases} \Phi(x, t) & \text{si } t < T_1, \\ Z_1 & \text{sinon.} \end{cases}$$

où la variable aléatoire Z_1 est définie via le noyau de transition Q : pour toute fonction mesurable bornée φ ,

$$\mathbb{E} [\varphi(Z_1) \mid T_1, X_0 = x] = \int \varphi(u) Q(\Phi(x, T_1), du).$$

Partant de la position X_{T_1} , le temps de parcours $S_2 = T_2 - T_1$ et la future position Z_2 sont obtenus de la même manière, et ainsi de suite. On obtient un processus à temps continu (X_t) ponctué par des sauts ayant lieu aux instants T_n . Les temps de parcours du flot sont les $S_n = T_n - T_{n-1}$. Enfin, les positions lors des sauts sont les $Z_n = X_{T_n}$.

Il est important de remarquer que la suite (Z_n) forme une chaîne de Markov. L'hypothèse principale que nous faisons dans la suite concerne l'ergodicité de ce processus. Nous imposons également des conditions de régularité des différentes caractéristiques du modèle. Le noyau Q est supposé être à densité par rapport à la mesure de Lebesgue. Enfin, une hypothèse de mélange de la chaîne de Markov (Z_n) assure la normalité asymptotique de l'estimateur. Nous encourageons le lecteur à se référer à [1] pour de plus amples détails.

Nous nous intéressons désormais à l'estimation du taux λ à partir de l'observation en temps long d'une seule trajectoire du processus lors des instants de saut. Nous avons donc accès à la suite des positions Z_n et aux temps de parcours S_n .

2 Procédure d'estimation

La fonction composée $\lambda \circ \Phi(x, t)$ décrit le taux d'apparition de la durée S_{n+1} en t partant de $Z_n = x$. C'est à cette quantité que nous avons facilement accès. Nous proposons de l'estimer ainsi

$$\widehat{\lambda \circ \Phi}^n(x, t) = \frac{\sum_{i=0}^{n-1} \frac{1}{v_i^d w_i} \mathbb{K}_d\left(\frac{Z_i - x}{v_i}\right) \mathbb{K}_1\left(\frac{S_{i+1} - t}{w_i}\right)}{\sum_{i=0}^{n-1} \frac{1}{v_i^d} \mathbb{K}_d\left(\frac{Z_i - x}{v_i}\right) \mathbb{I}_{\{S_{i+1} > t\}}},$$

où \mathbb{K}_p est une fonction noyau sur \mathbb{R}^p , et où les fenêtres sont définies pour tout entier k par $v_k = v_0(k+1)^{-\alpha}$ et $w_k = w_0(k+1)^{-\beta}$. Les noyaux \mathbb{K}_1 et \mathbb{K}_d doivent satisfaire un certain nombre d'hypothèses de régularité que nous ne détaillons pas ici. Les paramètres α et β admissibles sont donnés par

$$\mathcal{A} = \{(\alpha, \beta) \in \mathbb{R}^2 : \alpha > 0, \beta > 0, \alpha d + \beta < 1, \alpha d + \beta + 2 \min(\alpha, \beta) > 1\}.$$

Notre résultat principal, obtenu via la théorie des martingales vectorielles, est le suivant.

Théorème. Pour tout couple (x, t) chargé par la loi invariante de (Z_n, S_{n+1}) , pour tout $(\alpha, \beta) \in \mathcal{A}$, il existe des fenêtres initiales v_0 et w_0 telles que

$$\widehat{\lambda \circ \Phi}^n(x, t) \xrightarrow{p.s.} \lambda \circ \Phi(x, t),$$

et

$$n^{\frac{1-\alpha d - \beta}{2}} \left(\widehat{\lambda \circ \Phi}^n(x, t) - \lambda \circ \Phi(x, t) \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{\tau_1^2 \tau_d^2 \lambda \circ \Phi(x, t)}{(1 + \alpha d + \beta) \Gamma(x, t)} \right)$$

où $\tau_p^2 = \int_{\mathbb{R}^p} \mathbb{K}_p^2 d\lambda_p$.

Ce résultat est intéressant mais ne répond pas pleinement à la question de l'estimation du taux $\lambda(x)$ en un point $x \in E$. Soit \mathcal{C}_x la courbe décrite par le flot parcouru à rebours partant de x ,

$$\mathcal{C}_x = \{\Phi(x, -t) : 0 \leq t < t^-(x)\}.$$

Alors, pour tout élément ξ de la courbe, il existe un unique temps $\tau_x(\xi)$ tel que

$$\Phi(\xi, \tau_x(\xi)) = x.$$

Par conséquent, quelque soit le point $\xi \in \mathcal{C}_x$, on a

$$\lambda \circ \Phi(\xi, \tau_x(\xi)) = \lambda(x).$$

Cette remarque nous permet de construire une famille d'estimateurs $\widehat{\lambda}_\xi^n(x)$ de $\lambda(x)$ indicés par les éléments ξ de \mathcal{C}_x ,

$$\forall \xi \in \mathcal{C}_x, \widehat{\lambda}_\xi^n(x) = \widehat{\lambda \circ \Phi}^n(\xi, \tau_x(\xi)).$$

De plus quelque soit l'indice ξ choisi, nous avons

$$\widehat{\lambda}_\xi^n(x) \xrightarrow{p.s.} \lambda(x) \quad \text{et} \quad n^{\frac{1-\alpha d-\beta}{2}} \left(\widehat{\lambda}_\xi^n(x) - \lambda(x) \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_\xi^2(x)),$$

où la variance asymptotique est donnée par

$$\sigma_\xi^2(x) = \frac{\tau_1^2 \tau_d^2 \lambda(x)}{(1 + \alpha d + \beta) \Gamma(\xi, \tau_x(\xi))}.$$

Nous choisissons donc d'approcher $\lambda(x)$ par l'estimateur $\widehat{\lambda}_\xi^n(x)$ de variance minimale. Cela revient à déterminer l'élément ξ de la courbe qui maximise la quantité $\Gamma(\xi, \tau_x(\xi))$. Malheureusement, cette fonction est inconnue et est donc à estimer. On propose de l'approcher par,

$$\forall \xi \in \mathcal{C}_x, \quad \widehat{\Gamma}^n(\xi, \tau_x(\xi)) = \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{v_i^d} \mathbb{K}_d \left(\frac{Z_i - x}{v_i} \right) \mathbb{I}_{\{S_{i+1} > \tau_x(\xi)\}}.$$

En conséquence, nous estimons le taux de saut $\lambda(x)$ ainsi,

$$\widehat{\lambda}_*^n(x) = \widehat{\lambda}_{\widehat{\xi}_*^n}^n(x), \quad \text{où} \quad \widehat{\xi}_*^n = \arg \max_{\xi \in \mathcal{C}_x} \widehat{\Gamma}_x^n(\xi, \tau_x(\xi)).$$

Comme souvent, les choix des paramètres de lissage α et β sont cruciaux, en particulier à l'étape de maximisation (en témoignent les fortes oscillations ou au contraire l'estimation trop lisse lorsque la fenêtre est mal choisie). Un critère classique pour choisir ce paramètre est de minimiser le critère ISE (Integrated Square Error en anglais). La validation croisée est une technique populaire pour sélectionner le paramètre minimisant ce critère. Nous soulignons ici deux difficultés : tout d'abord les estimateurs sont calculés à partir de données dépendantes. De plus, presque sûrement, il n'y a aucune donnée sur l'ensemble d'intégration du ISE. Nous renvoyons le lecteur à notre article [1] dans lequel nous présentons une méthode ad hoc pour résoudre ce problème.

3 Simulation

Dans cette partie, on considère un PDMP (X_t) à valeurs dans $E = (0, 1)^2$ (voir Figure 1). Le flot Φ est défini par, $\forall x = (x_1, x_2) \in \mathbb{R}^2$, $\forall t \in \mathbb{R}$, $\Phi(x, t) = (x_1 + t, x_2)$. Le taux de saut λ est défini par, $\forall x \in \mathbb{R}^2$, $\lambda(x) = x_1 + x_2$. Enfin, le noyau de transition Q est défini par, $\forall x \in E$, $\forall A \in \mathcal{B}(E)$, $Q(x, A) \propto \int_A u(1 - u)^{2/x_1 - 1} v(1 - v) du dv$. Partant de (x_1, x_2) , le processus évolue dans le carré unité, toujours vers la droite, jusqu'au premier instant de saut. Cet instant correspond soit à l'instant où le processus atteint le bord $\{(\xi_1, \xi_2) : \xi_1 = 1\}$ soit, si le processus saute avant d'atteindre le bord, à un instant aléatoire, de taux de saut égal à $x_1 + x_2 + t$, c'est-à-dire selon une loi de Weibull (voir

Figure 1). Ce modèle correspond à un processus de type TCP avec environnement utilisé pour décrire le transfert de données à travers le réseau Internet : la première composante modélise le nombre de paquets de données transmis par unité de temps (les instants de saut correspondent à des apparitions de congestion) alors que la seconde dimension modélise la qualité du réseau (la probabilité de congestion est d'autant plus grande que la seconde variable est élevée).

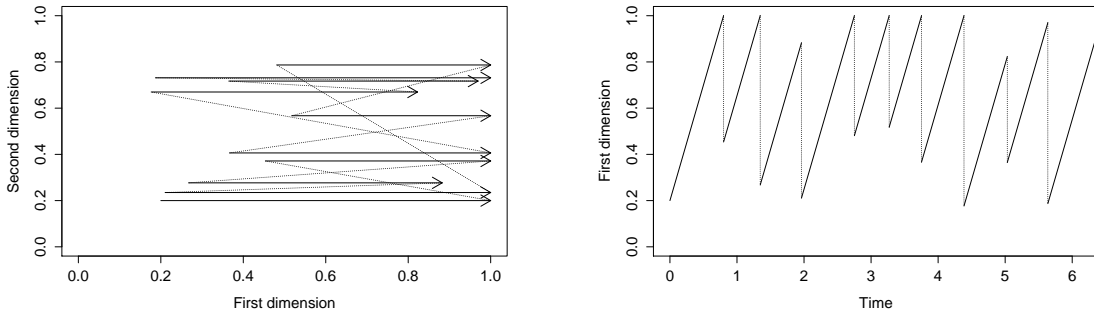


Figure 1: Deux représentations de la même trajectoire du processus simulé jusqu'au 10^{ème} saut.

Nous estimons le taux de saut λ au point $x = (0.75, 0.5)$. La classe d'estimateurs de $\lambda(x)$ est donc indexée par les éléments $\xi \in \mathcal{C}_x = (0, 0.75] \times \{0.5\}$. Nous observons la chaîne de Markov (Z_n, S_{n+1}) jusqu'au 10 000^{ème} saut. Nous présentons ci-dessous les résultats de nos simulations sur 100 répétitions de l'expérience (voir Figure 2).

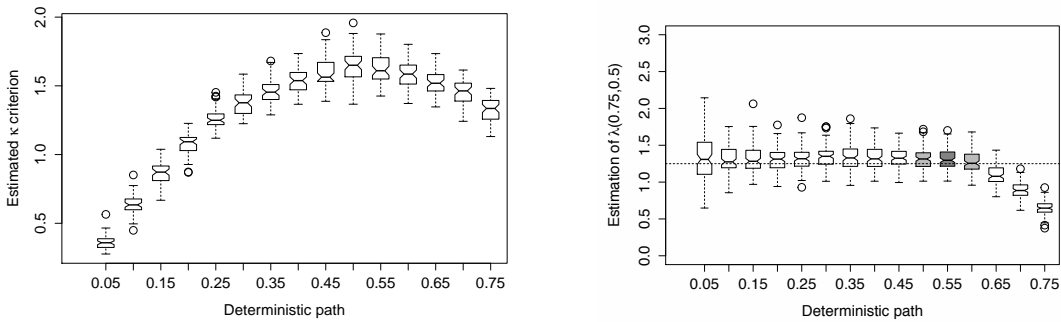


Figure 2: L'indice optimal $\hat{\xi}_*^n$ est calculé en maximisant le critère proposé (à gauche). L'estimateur $\hat{\lambda}_\xi^n(x)$ est calculé pour différentes valeurs de ξ (à droite). En couleur, les estimateurs sélectionnés $\xi = \hat{\xi}_*^n$ correspondent au meilleur choix à la fois en terme de biais et variance.

Bibliographie

- [1] R. Azaïs, A. Muller-Gueudin, *Optimal choice among a class of nonparametric estimators of the jump rate for piecewise-deterministic Markov processes*, (2015), Preprint hal-01168651.
- [2] P. Bertail, S. Clemençon, J. Tressou, *Statistical analysis of a dynamic model for dietary contaminant exposure*, (2010), Journal of Biological Dynamics, 4(2), pp 212-234.
- [3] J. Chiquet, N. Limnios, *A method to compute the transition function of a piecewise deterministic Markov process with application to reliability*, (2008), Statistics & Probability Letters, 78(12), pp1397-1403.
- [4] A. Crudu, A. Debussche, A. Muller, O. Radulescu, *Convergence of stochastic gene networks to hybrid piecewise deterministic processes*, (2012), The Annals of Applied Probability, 22(5).
- [5] M. H. A. Davis, *Markov models and optimization*, Chapman & Hall, 1993.
- [6] B. de Saporta, F. Dufour, H. Zhang, C. Elegbede, *Optimal stopping for the predictive maintenance of a structure subject to corrosion*, (2012) Journal of Risk and Reliability, 226(2), pp169-181.
- [7] A. Genadot, M. Thieullen, *Averaging for a fully coupled piecewise-deterministic Markov process in infinite dimensions*, (2012), Advances in Applied Probability, 44(3), pp 749-773.
- [8] O. Radulescu, A. Muller, A. Crudu, *Théorèmes limites pour les processus de Markov à sauts*, (2007), Technique et Science Informatiques, 26(3-4), pp443-469.
- [9] L. Robert, M. Hoffmann, N. Krell S. Aymerich, J. Robert, M. Doumic, *Division in Escherichia coli is triggered by a size-sensing rather than a timing mechanism*, (2014), BMC Biology, 12(1).